Online Transfer Learning for Differential Diagnosis of Benign and Malignant Thyroid Nodules With Ultrasound Images

Hui Zhou, Kun Wang, and Jie Tian, *Fellow, IEEE*

Abstract*—*Objective: **We aimed to propose a highly automatic and objective model named online transfer learning (OTL) for the differential diagnosis of benign and malignant thyroid nodules from ultrasound (US) images.** ***Methods:*** **The OTL mothed combined the strategy of transfer learning and online learning. Two datasets (1750 thyroid nodules with 1078 benign and 672 malignant nodules, and 3852 thyroid nodules with 3213 benign and 639 malignant nodules) were collected to develop the model. The diagnostic accuracy was also compared with VGG-16 based transfer learning model and different input images based model. Analysis of receiver operating characteristic (ROC) curves were performed to calculate optimal area under it (AUC) for benign and malignant nodules.** ***Results:*** **AUC, sensitivity and specificity of OTL were 0.98 (95% confidence interval [CI]: 0.97–0.99), 98.7% (95% confidence interval [CI]: 97.8%–99.6%) and 98.8% (95% confidence interval [CI]: 97.9%–99.7%) in the final online learning step, which was significantly better than other deep learning models ($P < 0.01$).** ***Conclusion:*** **OTL model shows the best overall performance comparing with other deep learning models. The model holds a good potential for improving the overall diagnostic efficacy in thyroid nodule US examinations.** ***Significance:*** **The proposed OTL model could be seamlessly integrated into the conventional work-flow of thyroid nodule US examinations.**

***Index Terms*—Diagnosis, online learning, radiomics, transfer learning, thyroid nodules, ultrasound images.**

Manuscript received November 21, 2019; revised January 8, 2020; accepted January 28, 2020. Date of publication February 3, 2020; date of current version September 18, 2020. This work was supported in part by Ministry of Science and Technology of China under Grant 2017YFA0205200, in part by the National Natural Science Foundation of China under Grants 81227901, 81527805, and 61671449, and in part by the Chinese Academy of Sciences under Grants GJJSTD20170004, KFJ-STS-ZDTP-059, 81930053, YJKYYQ2018 0048, and XDB32030200. *(Corresponding author: Jie Tian.)*

H. Zhou and K. Wang are with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences.

J. Tian is with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China and with the School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China and also with the Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, Beihang University, Beijing, China (e-mail: jie.tian@ia.ac.cn).

This article has supplementary downloadable material available at https://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TBME.2020.2971065

I. Introduction

THYROID NODULES are defined as discrete lesions within the thyroid gland, radiologically distinct from surrounding thyroid parenchyma [1]. They are becoming increasingly common in clinical practice, being detected in up to 65% of the general population in the world [2]. As the majority of thyroid nodules are benign or behave indolently [3], the accurate identification of benign and malignant thyroid nodules is vital in clinical decision-making and management.

In clinic, Fine-needle aspiration (FNA) biopsy has been treated as the golden standard for the diagnosis of benign and malignant thyroid nodules [1], [4]. However, it is invasive and limited by specimen collection and operator experience [5]. Currently, ultrasound (US) is the first clinical choice of thyroid nodules screening, because of its non-radioactivity, easy-to-operate, and rapid diagnostic work-up [6]. Therefore, US features can be utilized to differentiate malignancies from benign thyroid nodules [7], [8].

At the very beginning, some studies demonstrated that a combination of US features provided certain diagnostic accuracy [9]. However, many other studies indicated a considerable overlap of US features appearing in both benign and malignant nodules [10], [11]. The sensitivity and specificity of using US for thyroid cancer diagnosis varied from 27% to 63% and 78.0% to 96.6% in various studies [1], [8] and [12]. This is likely due to interobserver variability in assigning sonographic features to nodules and that US is highly operator dependent.

In contrast, an emerging technology named radiomics based on machine learning can extract and analyze thousands of quantitatively calculated image features (also called radiomics features) from medical images, which has the potential to uncover disease characteristics that is impossible for human to recognize by naked eyes in clinical practice [13]. Radiomics has been proven useful for analyzing magnetic resonance imaging (MR) and computed tomography (CT) images with impressive effectiveness [14]–[17], but its applications in US are still rarely reported [9], [18]–[22]. Therefore, it is worthy of investigating whether a radiomics approach can make better use of thyroid ultrasound images.

There are only a few radiomics studies that have been conducted on ultrasound images for classifying benign and malignant thyroid nodules [9], [18] and [21]. However, they were limited either by the relatively small number of patient population

TABLE I
THE DISTRIBUTION OF BENIGN AND MALIGNANT THYROID NODULES IN TRAINING AND VALIDATION COHORT FOR DATASET 1

Cohorts	Total	Benign nodules	Malignant nodules
Total	1750	1078	672
Training	1167	719	448
Validation	583	359	224

[9], [18], or lacking cytology results as gold standard [18], or too much labor work for operators [9], [21].

Here, we developed an online transfer learning (OTL) method based on transfer learning and online learning strategy for the quantitative analysis of thyroid ultrasound images. It is a deep learning based radiomics approach, and does not require complicated manual segmentations of thyroid nodule boundaries. Specifically, our contributions are threefold.

1) We built a deep learning based radiomics method for differential diagnosis of benign and malignant thyroid nodules in ultrasound images. This method can achieve automatic feature extraction, these meaningful features include high-level and high-abstract information that allow us to get a good classification accuracy.
2) The proposed OTL model adopted transfer learning strategy, which reused the features extracted from pre-trained model, and we further compared the accuracy with different pre-trained models. To the best of our knowledge, it is the first attempt that reuses a model pre-trained by ultrasound images. Our method showed better results from US based pre-trained radiomics model.
3) The proposed OTL model adopted online learning strategy, which made it easier to improve our model through continuous medical images, since the images from clinic usually come as a batch. The OTL method could be seamlessly integrated into the conventional work-flow of clinical thyroid US examinations.

II. MATERIAL AND METHODS

A. Datasets

The thyroid images used in our research were from two datasets, one is from Ningbo No.2 hospital, called dataset 1, and another is from a public available dataset proposed by Pedraza [23], called dataset 2. This retrospective study was approved by the Ethics Committee of Ningbo No.2 hospital in China.

Dataset 1 consists of 1750 thyroid nodules, of which 1078 nodules are labeled as benign and 672 nodules are labeled as malignant according to their FNA results. Dataset 2 consists of 3852 thyroid nodules, of which 3213 nodules are labeled as benign and 639 nodules are labeled as malignant confirmed by FNA results. Their distribution are summarized in Table I, Table II and Table III. Examples of benign and malignant thyroid nodules ultrasound images for two datasets are displayed in Fig. 1.

B. Preprocessing

For applying OTL, we designed a simple manual initiation by defining region-of-interests (ROIs). For each thyroid nodule, one

TABLE II
THE DISTRIBUTION OF BENIGN AND MALIGNANT THYROID NODULES IN TRAINING AND VALIDATION COHORT FOR PART 1 OF DATASET 2

Cohorts	Total	Benign nodules	Malignant nodules
Total	963	803	160
Training	642	535	107
Validation	321	268	53

TABLE III
THE DISTRIBUTION OF BENIGN AND MALIGNANT THYROID NODULES IN DIFFERENT ONLINE LEARNING STEP FOR PART 2 OF DATASET 2

Steps	Total	Benign nodules	Malignant nodules
Total	2889	2407	482
Step 1	963	803	160
Step 2	963	802	161
Step 3	963	802	161

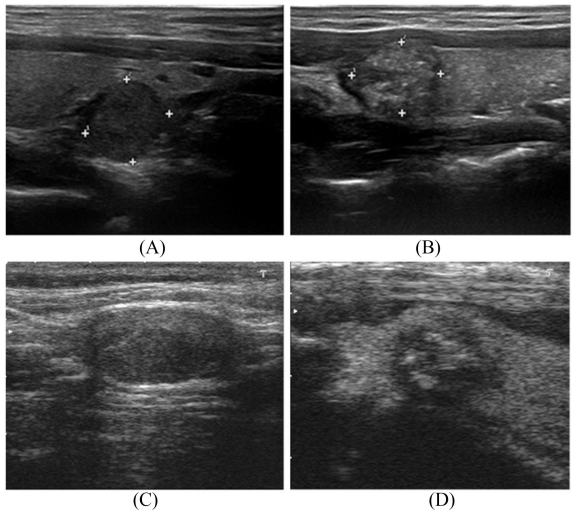

Fig. 1. Examples of benign and malignant thyroid nodules ultrasound images. (A) and (B) benign and malignant thyroid nodules from dataset1, (C) and (D) benign and malignant thyroid nodules from dataset 2.

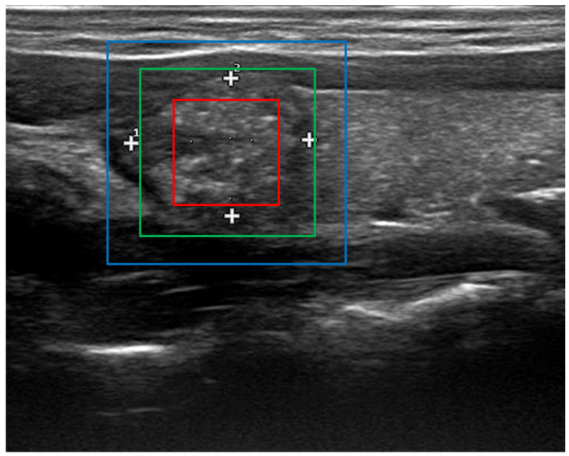

Fig. 2. Illustration of the region-of-interests (ROIs). Three different size of ROIs (sizes: 150 × 150 pixels for the red, 200 × 200 pixels for the green and 250 × 250 pixels for the blue) were automatically generated by a simple designed manual initiation.

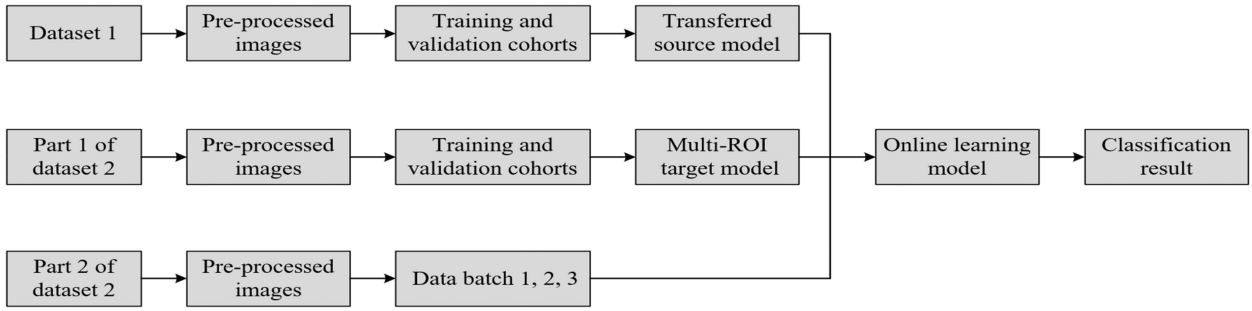

Fig. 3. Illustration of online transfer learning (OTL) flowchart. Dataset 1 and part 1 of dataset 2 were used to develop the transferred source model and Multi-ROI target model, respectively. Part 2 of dataset 2 was used for online learning procedure.

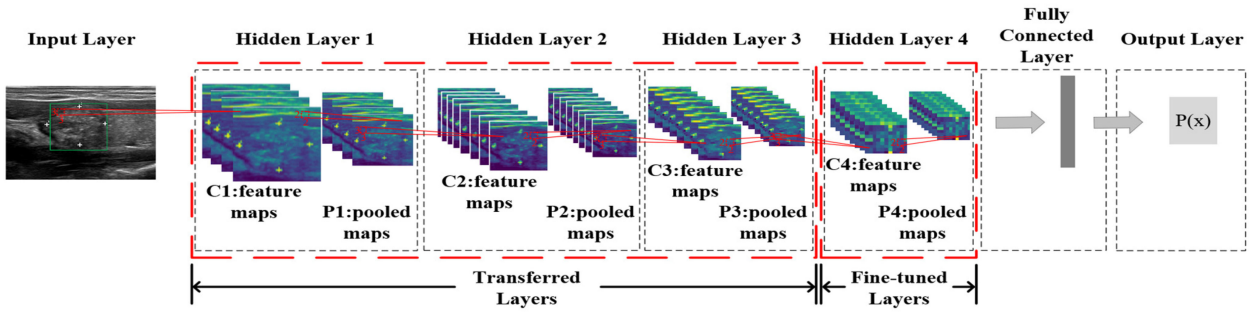

Fig. 4. Illustration of source domain model flowchart. The region-of-interests (ROI) was sent into the input layer, followed by three transferred layers, a fine-tuned layer was connected to the transferred layers, and then a fully connected layer was connected to the fine-tuned layer to combine different features extracted by the previous layers, at last an output layer was used to calculate the probability for the classification.

square ROI (size: 200 × 200 pixels) and three square ROIs (sizes: 150 × 150 pixels, 200 × 200 pixels and 250 × 250 pixels) were automatically generated after one mouse click on the nodule center area (Fig. 2) for dataset 1 and dataset 2, respectively. These sizes of ROIs were based on the statistical analysis of the average size of thyroid nodules, and three targeting regions were roughly inside, around and outside the thyroid nodule, respectively.

To reduce the potential bias caused by the unbalanced data and the limited size of population, we applied the data augmentation strategy [24] including the rotation, shift and flip operation. Thyroid images were augmented through a number of random transformations, which increased the data pool and decreased the overfitting of the generated radiomics model.

C. Online Transfer Learning

The proposed OTL model was combined with two major strategies, one was transfer learning and another was online learning (Fig. 3). First we used transfer learning to train our source domain model on dataset 1, then we used multi-ROI input strategy to train our target domain model on part 1 of dataset 2, finally we adopted online transfer learning to combine the two models and test it on part 2 of dataset 2 with different batch of images.

1) Source Domain Model: To develop the source domain model, enrolled patients from dataset 1 were randomly divided into the training cohort and validation cohort (Table I). US images and FNA biopsies of the training cohort were used to optimize a large number of parameters in the CNN model, whereas the validation cohort was to evaluate the performance of the trained model. The transfer learning was based on convolutional neural network (CNN) architecture. It consisted of four hidden layers. The first three layers were transferred from one of our previous studies without any modification [25], whereas the last hidden layer was fine-tuned using enrolled thyroid US images. This layer contained 32 feature maps, and the size of the convolution filter and the max pooling was 3 × 3 pixels and 2 × 2 pixels, respectively. Finally, a fully-connected layer with 32 nodes was connected to every neuron in the last three pooling layers, and the probability (a malignancy score) of the binary classification (benign or malignant) can be calculated in the output layer (Fig. 4). We also compared the classification accuracy between our previous model which was pre-trained by ultrasound images and VGG-16 model which was pre-trained by natural images.

2) Target Domain Model: To develop the target domain model, enrolled patients from part 1 of dataset 2 were randomly divided into the training cohort and validation cohort (Table II). The target domain model was the same architecture as source domain model, except for the input layer and we did not use the transfer learning strategy here. For input layer, three cropped images were used, and their information were combined at the fully connected layer (Fig. 5). We also compared the classification accuracy by using only one ROI (size: 200 × 200 pixels) and all three ROIs (sizes: 150 × 150 pixels, 200 × 200 pixels and 250 × 250 pixels).

3) OTL Model: To develop OTL model, we adopted online learning strategy [26], which aimed to attack an online learning

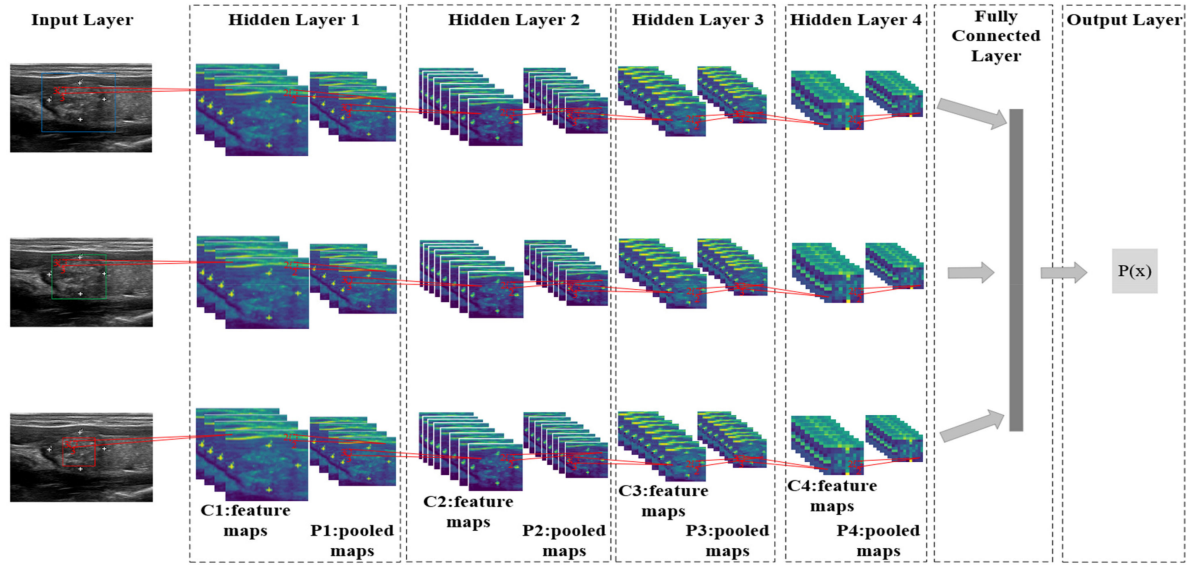

Fig. 5. Illustration of target domain model flowchart. Three different size of region-of-interests (ROIs) were sent into the input layer, followed by four hidden layers, and then a fully connected layer was connected to the hidden layer to combine different features extracted by the previous layers, at last an output layer was used to calculate the probability for the classification.

task on a target domain by transferring knowledge from some source domain. Here, OTL model learned an ensemble prediction function that is the mixture of both the source and the target prediction functions, i.e., $\mathrm{h}(\mathrm{x})$ and $\mathrm{f}(\mathrm{x})$, which x stood for the input image. In order to effectively combine the two prediction functions $\mathrm{h}(\mathrm{x})$ and $\mathrm{f}(\mathrm{x})$ at the t-trial of the online learning task, we introduced two combination weighting parameters, $w_{1,t}$ and $\mathrm{w}_{2,t}$, for the two prediction functions respectively. At the t-th step, given an instance $\mathrm{x}_{2,t}$ from part 2 of dataset 2 (Table III), we predicted its class label by the following prediction function:

$$\tilde{y}_t = \mathrm{sign}\left(w_{1,t}\mathrm{h}\left(\mathrm{x}_{2,t}\right) + \mathrm{w}_{2,t}\mathrm{f}\left(\mathrm{x}_{2,t}\right)\right) \tag{1}$$

At the beginning of the OTL task, we simply initialized $w_{1,1} = \mathrm{w}_{2,1} = \frac{1}{2}$. In order to updating the weighting parameters $w_{1,t}$ and $\mathrm{w}_{2,t}$, we adopted the following updating function:

$$w_{1,t+1} = \frac{w_{1,t} * s_t\left(h\right)}{w_{1,t} * s_t\left(h\right) + w_{2,t} * s_t\left(f_t\right)} \tag{2}$$

$$w_{2,t+1} = \frac{w_{2,t} * s_t\left(f_t\right)}{w_{1,t} * s_t\left(h\right) + w_{2,t} * s_t\left(f_t\right)} \tag{3}$$

where $s_t(u) = \exp\{-nl(u(\mathrm{x}_{2,t}), \mathrm{y}_{2,t})\}$, $l(z, y)$ is a loss function, and $l(z, y) = (z - y)^2$. Finally, Fig. 6 summarized the proposed online transfer learning strategy.

The detailed introduction and mathematical descriptions of CNN model, transfer learning strategy, VGG-16 model and Multi-ROI input strategy were demonstrated in supplementary materials.

D. Statistical Analysis

Descriptive statistics were summarized as 95% confidence interval (CI). Analysis of receiver operating characteristic (ROC) curves were performed to calculate optimal area under it (AUC)

Input: the transferred source model $\mathrm{h}(\mathrm{x})$ and multi-ROI target model $\mathrm{f}(\mathrm{x})$

Initial $w_{1,1} = \mathrm{w}_{2,1} = \frac{1}{2}$

for t = 1, 2, 3 **do**

receive data batch $\mathrm{x}_{2,t}$

predict $\tilde{y}_t = \mathrm{sign}(w_{1,t}\mathrm{h}(\mathrm{x}_{2,t}) + \mathrm{w}_{2,t}\mathrm{f}(\mathrm{x}_{2,t}))$

receive correct label {-1, +1}

compute $w_{1,t+1} = \frac{w_{1,t} * s_t(h)}{w_{1,t} * s_t(h) + w_{2,t} * s_t(f_t)}$, $w_{2,t+1} = \frac{w_{2,t} * s_t(f_t)}{w_{1,t} * s_t(h) + w_{2,t} * s_t(f_t)}$

end for

Fig. 6. Illustration of online transfer learning strategy. The weighting parameters $w_{1,t}$ and $\mathrm{w}_{2,t}$ were updated by three different batch of data from part 2 of dataset 2.

for benign and malignant nodules. Differences between various AUCs were compared by using a Delong test. Sensitivity, specificity were also calculated. P values less than 0.05 indicated statistical significance. The statistical analyses were performed using SPSS software for Windows, version 20.0 (SPSS, Chicago, IL).

III. Results

A. Diagnostic Accuracy of Source Domain Model

Both in training and validation cohorts from dataset 1, US based transfer learning model demonstrated higher diagnostic accuracy comparing with the VGG-16 based transfer learning model for the differential diagnosis of benign and malignant thyroid nodules (Fig. 7). Differences of AUCs were all statistically significant ($P < 0.05$, Table IV). AUCs of US based transfer learning model reached 0.91 and 0.90 in training and validation cohorts, respectively, which were 0.05 and 0.05 higher than these of the VGG-16 based transfer learning model.

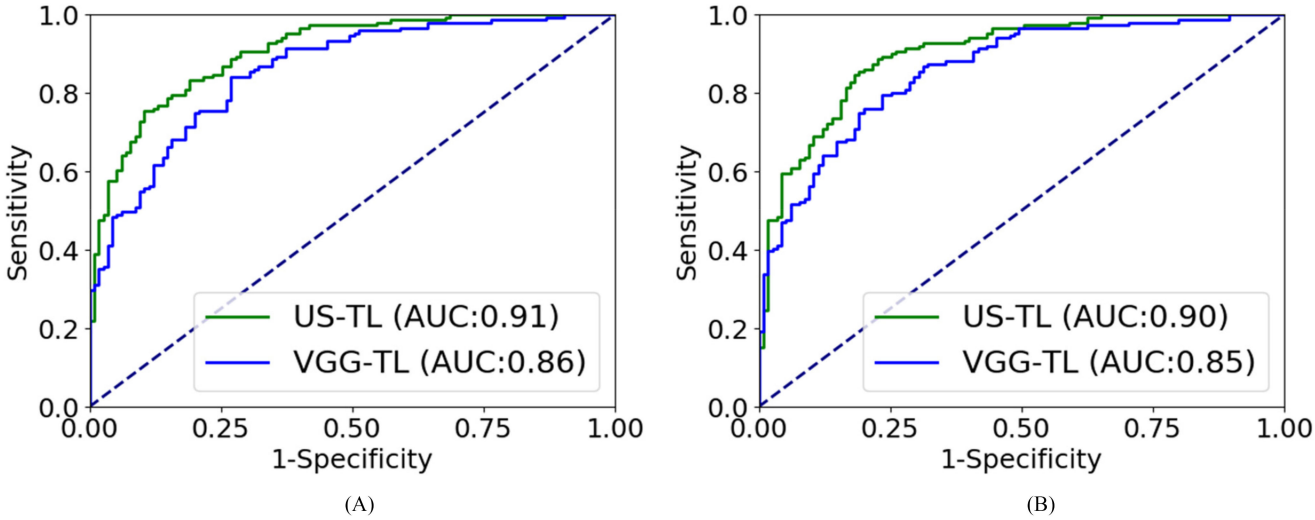

Fig. 7. Comparison of receiver operating characteristic (ROC) curves, area under the curve (AUC) between transferred models (US-TL and VGG-TL) for the differential diagnosis of benign and malignant thyroid nodules in training and validation cohorts, respectively. (A) and (B) ROC curves in training and validation cohorts, respectively.

TABLE IV
DIAGNOSTIC PERFORMANCE OF US BASED AND VGG-16 BASED TRANSFER LEARNING FOR THE DIFFERENTIAL DIAGNOSIS OF BENIGN AND MALIGNANT THYROID NODULES IN TRAINING VALIDATION COHORTS

Models	Cohorts	AUC	Sensitivity %	Specificity %
US-TL	T	0.91 (0.89-0.93)	83.1 (81.6-84.6)	80.2 (78.5-81.9)
	V	0.90 (0.88-0.92)	82.3 (80.8-83.8)	81.1 (79.4-82.8)
VGG-TL	T	0.86* (0.84-0.88)	74.6 (72.8-76.4)	79.8 (78.1-81.5)
	V	0.85* (0.83-0.87)	76.7 (74.9-78.5)	79.5 (77.6-81.4)

Statistical quantifications were demonstrated with 95% confidence interval.
Abbreviations: T, training cohort; V, validation cohort; AUC, area under the receiver-operator-characteristic curve; US-TL, US based transfer learning; VGG-TL, VGG-16 based transfer learning.
AUC of US-TL was statistically compared to AUC of VGG-TL (*$P < 0.05$).

B. Diagnostic Accuracy of Target Domain Model

Both in training and validation cohorts from part 1 of dataset 2, three ROIs based CNN model demonstrated higher diagnostic accuracy comparing with one ROI based model for the differential diagnosis of benign and malignant thyroid nodules (Fig. 8). Differences of AUCs were all statistically significant ($P < 0.05$, Table V). AUCs of three ROIs based model reached 0.88 and 0.87 in training and validation cohorts, respectively, which were 0.05 and 0.05 higher than these of one ROI based model.

C. Diagnostic Accuracy of OTL Model

As US based transfer learning model and three ROIs based model showed the best performance over other approaches, we further investigated whether its diagnostic accuracy was influenced by different batch of images by using online learning in part 2 of dataset 2 (Fig. 9). Differences of AUCs were all statistically significant ($P < 0.05$, Table VI). AUCs reached 0.92, 0.95 and 0.98 at different steps.

IV. DISCUSSION

In this study, we developed an online transfer learning model called OTL for the differential diagnosis of benign and malignant thyroid nodules by automatic and quantitative analysis of thyroid US images. We retrospectively enrolled US images and FNA results from two datasets. Dataset 1 and part 1 of dataset 2 were used to develop source and target domain model, respectively. Part 2 of dataset 2 was used to test the performance of the proposed OTL model.

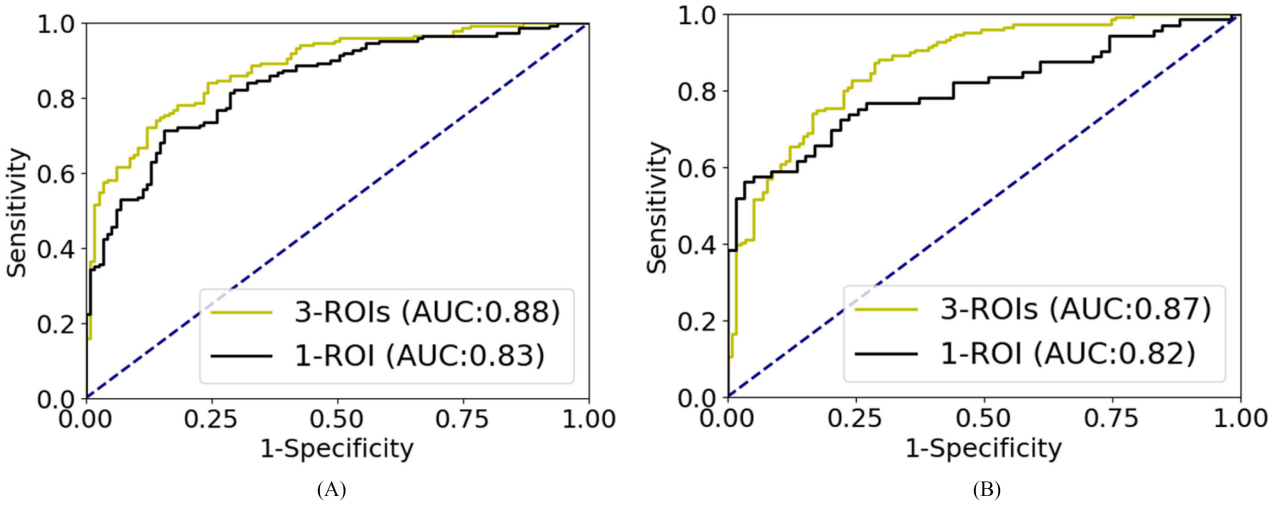

Fig. 8. Comparison of receiver operating characteristic (ROC) curves, area under the curve (AUC) between different inputs models (3-ROIs and 1-ROI) for the differential diagnosis of benign and malignant thyroid nodules in training and validation cohorts, respectively. (A) and (B) ROC curves in training and validation cohorts, respectively.

TABLE V
DIAGNOSTIC PERFORMANCE OF THREE ROIS AND ONE ROI FOR THE DIFFERENTIAL DIAGNOSIS OF BENIGN AND MALIGNANT THYROID NODULES IN TRAINING VALIDATION COHORTS

Inputs	Cohorts	AUC	Sensitivity %	Specificity %
3 ROIs	T	0.88 (0.86-0.90)	79.1 (77.6-80.6)	78.2 (76.5-79.9)
	V	0.87 (0.85-0.89)	77.9 (76.1-79.7)	77.8 (76.1-79.5)
1 ROI	T	0.83* (0.81-0.85)	72.7 (70.9-74.5)	76.3 (74.8-77.8)
	V	0.82* (0.80-0.84)	73.2 (71.3-75.1)	76.1 (74.3-77.9)

Statistical quantifications were demonstrated with 95% confidence interval.
Abbreviations: T, training cohort; V, validation cohort; AUC, area under the receiver-operator-characteristic curve.
AUC of three ROIs was statistically compared to AUC of one ROI ($^*P < 0.05$).

TABLE VI
DIAGNOSTIC PERFORMANCE OF DIFFERENT STEPS OF ONLINE LEARNING FOR THE DIFFERENTIAL DIAGNOSIS OF BENIGN AND MALIGNANT THYROID NODULES

Steps	AUC	Sensitivity %	Specificity %
Step 1	0.92* (0.91-0.93)	85.1 (83.6-86.6)	85.3 (83.5-87.1)
Step 2	0.95* (0.94-0.96)	90.3 (88.6-92.0)	89.9 (88.1-91.6)
Step 3	0.98 (0.97-0.99)	98.7 (97.8-99.6)	98.8 (97.9-99.7)

Statistical quantifications were demonstrated with 95% confidence interval.
Abbreviations: AUC, area under the receiver-operator-characteristic curve.
AUC of step 3 was statistically compared to AUC of step 1 and step 2, respectively ($^*P < 0.05$).

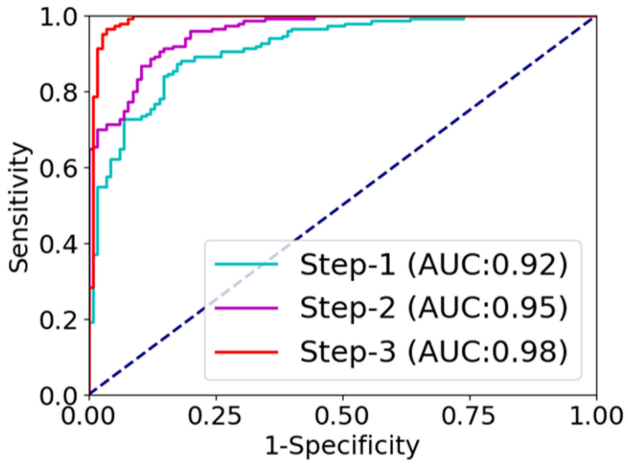

Fig. 9. Comparison of receiver operating characteristic (ROC) curves, area under the curve (AUC) between different steps of online learning produce (step 1, step 2 and step 3) for the differential diagnosis of benign and malignant thyroid nodules in three different batch of data, respectively.

The comparison experiment for transfer learning strategy demonstrated that US based transfer learning model showed higher diagnostic accuracy comparing with VGG-16 based transfer learning. The results indicated that the US based strategy was useful for accuracy improvement, maybe because it was transferred from the same image domain, i.e., ultrasound images. The result also indicated that the US based transfer learning model is of great clinical value, since the transferred model was designed for US images on purpose.

The proposed multi-ROI model also indicated that the multiple ROIs strategy made critical contribution for accuracy improvement, because it enabled independent analysis targeting regions inside and outside each thyroid nodule. This strategy suggested that a combination of different region of thyroid nodules was critical for radiomics feature extraction procedure.

The OTL model combined transfer learning and online learning strategy, which demonstrated the highest diagnosis accuracy. It made full use of the features transferred from US based pre-trained model, as well as the advantages of online learning. Since the images from hospitals were collected over time, we could not process all the data at a time. The OTL model made it more realistic for clinical use due to its advantages in using data.

Our work demonstrated several advantages over other studies attempted to differentiate malignant and benign thyroid nodules through computer-aided analysis on thyroid US images [9], [18] and [21]. The majority of those studies used human-defined US features extracted from US images by a lot of labor work. OTL was a highly automatic end-to-end approach. It automatically extracted thousands of features from the images and adopted transfer learning and online learning strategy to optimize the diagnosis model. Besides that, compared with the basic CNN model [22] which used all the image data at the first time, the procedures of using OTL were highly consistent with the existing clinical data collection methods, so it could be seamlessly integrated into the conventional work-flow of thyroid US examinations without extra time and labor cost.

The major limitation in our study was that the data came from retrospective datasets. The performance of OTL needs to be further validated in a perspective study. A larger dataset acquired from different hospitals and different counties is necessary for consisting a more comprehensive dataset, so that the accuracy and reliability of OTL can be continuously improved.

V. Conclusion

In conclusion, OTL achieved the most accurate differential diagnosis of benign and malignant thyroid nodules comparing with transfer learning and multi-ROI based model. Its performance continuously improved with the increase of data. The model holds a good potential for improving the overall diagnostic efficacy in thyroid US examinations, and it is very consistent with the clinical operation procedure.

Acknowledgment

We would like to thank Doctor Baowen Zheng and Congde Chen for their help with the pathological diagnoses.

References

[1] B. R. Haugenet *et al.*, "2015 American Thyroid Association Management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer," *Thyroid*, vol. 26, no. 1, pp. 1–133, 2015.
[2] G. Russ *et al.*, "Thyroid incidentalomas: Epidemiology, risk stratification with ultrasound and workup," *Eur Thyroid J.*, vol. 3, no. 3, pp. 154–163, 2014.
[3] T. E. Angell *et al.*, "A cohort analysis of clinical and ultrasound variables predicting cancer risk in 20,001 consecutive thyroid nodules," *J. Clin. Endocrinol. Metab.*, vol. 104, no. 11, pp. 5665–5672, 2019.
[4] S. Daniel *et al.*, "The impact of benign gene expression classifier test results on the endocrinologist–patient decision to operate on patients with thyroid nodules with indeterminate fine-needle aspiration cytopathology," *Thyroid*, vol. 22, no. 10, pp. 996–1001, 2012.
[5] W. J. Moon *et al.*, "Benign and malignant thyroid nodules: US differentiation–Multicenter retrospective study," *Radiology*, vol. 247, no. 3, pp. 762–770, 2008.
[6] D. S. Cooper *et al.*, "Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer," *Thyroid*, vol. 19, no. 11, pp. 1167–1214, 2009.
[7] H. J. Moon *et al.*, "Diagnostic performance of gray-scale US and elastography in solid thyroid nodules," *Radiology*, vol. 262, no. 3, pp. 1002–1013, 2012.
[8] L. R. Remonti *et al.*, "Thyroid ultrasound features and risk of carcinoma: A systematic review and meta-analysis of observational studies," *Thyroid*, vol. 25, no. 5, pp. 538–550, 2015.
[9] S. Y. Kim *et al.*, "Application of texture analysis in the differential diagnosis of benign and malignant thyroid nodules: Comparison with gray-scale ultrasound and elastography," *AJR Amer. J. Roentgenol*. vol. 205, no. 3, pp. W343–351, 2015.
[10] J. D. Iannuccilli, J. J. Cronan, and J. M. Monchik "Risk for malignancy of thyroid nodules as assessed by sonographic criteria: The need for biopsy," *J. Ultrasound Med.*, vol. 23, no. 11, pp. 1455–1464, 2004.
[11] J. R. Wienke *et al.*, "Sonographic features of benign thyroid nodules: interobserver reliability and overlap with malignancy," *J. Ultrasound Med.*, vol. 22, no. 10, pp. 1027–1031, 2003.
[12] M. C. Frates *et al.*, "Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement," *Radiology*, vol. 237, no. 3, pp. 794–800, 2005.
[13] H. J. Aerts *et al.*, "Decoding tumor phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.*, vol. 5, 2014, Art. no. 4006.
[14] R. J. Gillies, P. E. Kinahan, and H. Hricak "Radiomics: Images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.
[15] Y. Q. Huang *et al.*, "Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer," *J. Clin. Oncol.*, vol. 34, no. 18, pp. 2157–2164, 2016.

[16] J. Shu *et al.*, "Clear cell renal cell carcinoma: CT-based radiomics features for the prediction of Fuhrman grade," *Eur. J. Radiol.*, vol. 109, pp. 8–12, 2018.
[17] X. Min *et al.*, "Multi-parametric MRI-based radiomics signature for discriminating between clinically significant and insignificant prostate cancer: Cross-validation of a machine learning method," *Eur. J. Radiol.*, vol. 115, pp. 16–21, 2019.
[18] J. Chi *et al.*, "Thyroid nodule classification in ultrasound images by fine-tuning deep convolutional neural network," *J Digit Imag.*, vol. 30, no. 4, pp. 477–486, 2017.
[19] J. Jiang *et al.*, "Computer aided diagnosis of Lymphoma based on dual-mode ultrasound radiomics," *Ultrasound Med. Biol.*, vol. 45, no. S18, 2019.
[20] Q. Zhang *et al.*, "Sonoelastomics for breast tumor classification: A radiomics approach with clustering-based feature selection on sonoelastography," *Ultrasound Med. Biol.*, vol. 43, no. 5, pp. 1058–1069, 2017.
[21] B. Zhang *et al.*, "Machine learning-assisted system for thyroid nodule diagnosis," *Thyroid*, vol. 29, no. 6, pp. 858–867, 2019.
[22] Y. Zhou *et al.*, "A radiomics approach with CNN for shear-wave elastography breast tumor classification," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 9, pp. 1935–1942, 2018.
[23] L. Pedraza *et al.*, "An open access thyroid ultrasound image database," vol. 9287, pp. w1–w6, 2015.
[24] V. Nitesh *et al.*, "SMOte: Synthetic minority over-sampling technique," *J Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
[25] K. Wang *et al.*, "Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: A prospective multicentre study," *Gut*, vol. 68, no. 4, pp. 729–741, 2018.
[26] P. Zhao *et al.*, "Online transfer learning," *Artif. Intell.*, vol. 216, pp. 76–102, 2014.